GENOME SEQUENCE, FUNCTIONAL GENOMICS OF SHIITAKE MUSHROOM *LENTINULA EDODES*

KWAN HOI SHAN, AU C.H., WONG M.C., QIN J., KWOK I.S.W., CHUM W.W.Y., YIP P.Y., WONG K.S., LI L., HUANG, Q.L.
School of Life Sciences and Food Research Centre, The Chinese University of Hong Kong
The Chinese University of Hong Kong, Shatin, N.T.
Hong Kong SAR, PRC
hoishankwan@cuhk.edu.hk

ABSTRACT

Lentinula edodes (Shiitake/Xianggu) is a popular cultivated mushroom species. Understanding the genomics and functional genomics of *L. edodes* is essential to improve its cultivation and quality. Genome sequencing of *L. edodes* provides numerous molecular genetic markers for breeding and genetic manipulation. We sequenced the genome of *L. edodes* monokaryon L54A using Roche 454 and ABI SOLiD. Sequencing reads of about 1011 Mb were *de novo* assembled into a 39.8 Mb genome. We compiled the genome sequences into a searchable database with which we have been annotating the genes and analyzing the metabolic pathways. Over 13,000 gene models were predicted from the genome sequence. The gene models were annotated by BLASTX and categorized according to Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG). For functional genomics, we have been using many molecular techniques including RNA arbitrarily primed-PCR, SAGE, LongSAGE, EST sequencing and cDNA microarray to analyze genes differentially expressed during development. Protein families of *L. edodes* genome sequence compared across genomes of several fungi identified protein families conserved to mushroom-forming fungi. We are learning more about the molecular biology and genetics of this economically important mushroom.

Keywords: Shiitake Mushroom, genome sequence, transcriptome, fruiting body development

INTRODUCTION

Lentinula edodes (Berk.) Pegler, or the shiitake mushroom, is the second most cultivated mushroom worldwide, especially in China and Japan. Understanding the genomics and functional genomics of *L. edodes* is essential to improve its cultivation and quality. *Lentinula edodes* follows a typical basidiomycete life cycle. Two monokaryotic mycelia with compatible mating types fuse to form a dikaryon. Under appropriate environmental conditions, dikaryotic mycelia aggregate to form a primordium and then mature into a fruiting body. The molecular biology of Agaricales fruiting body initiation and development remains to be elucidated.

Since 2008, reports of genome sequences from mushroom-forming basidiomycete fungi (Agaricales) has been emerging, including *Laccaria bicolor* [1], *Coprinopsis cinerea* [2], *Schizophyllum commune* [3]. Comparison of the mushroom genome sequences will help distinguish genes shared by different mushroom species from those that are specific to individual species. Further characterization of the conserved genes may provide insights into the complex developmental process.

We sequenced the genome of the monokaryotic *L. edodes* strain L54A. This provides resources for the study of growth and development of this mushroom. We compiled the genome sequences, annotations, genetic variations and mushroom genome comparisons into a genome database and analysis platform.

MATERIALS AND METHODS

Genomic DNA was obtained from *L. edodes* monokaryon L54A and subject to shotgun and paired-end sequencing using Roche 454 GS-FLX/Titanium system. The pool of 454 sequencing reads was *de novo* assembled using Newbler version 2.3 (Roche). The intrinsic sequence error of 454 sequencing associated with homopolymers was fixed by additional shotgun sequencing by ABI SOLiD 3 system.

Protein-coding genes were predicted using AUGUSTUS *L. edodes* model [4] (trained with *Schizophyllum commune* proteome, assisted with *L. edodes* ESTs). The reference gene set was subject to automated annotation based on the amino acid sequences. Protein domains and functional sites were predicted by InterProScan [5]. KEGG orthology identifiers were assigned and biological pathways were associated by KEGG Automatic Annotation Server (KAAS) [6]. Gene Ontology (GO) terms were assigned to genes by inferring GO terms of assigned Pfam protein domains [7].

Genome sequences and protein sequences of 14 other fungi, including 3 Agaricales, 4 other basidiomycetes and 7 ascomycetes (Table 1), were collected from public sequence databases. Protein sequences of all 15 fungal genomes were filtered, pooled together and subject to all-versus-all comparison by BLASTP [8]. The software OrthoMCL [9] was used to cluster the proteins into families based on the comparison results.

Table 1: List of 15 fungal genomes compared

Phylum	Species	Reference
Basidiomycota	**Agaricales**	
	Lentinula edodes	This study
	Laccaria bicolor	[1]
	Coprinopsis cinerea	[2]
	Schizophyllum commune	[3]
	Others	
	Cryptococcus neoformans	[10]
	Phanerochaete chrysosporium	[11]
	Postia placenta	[12]
	Ustilago maydis	[13]
Ascomycota	*Apergillus fumigates*	[14]
	Apergillus nidulans	[15]
	Apergillus oryzae	[16]
	Neurospora crassa	[17]
	Saccharomyces cerevisiae	[18]
	Trichoderma reesei	[19]
	Tuber melanosporum	[20]

RESULTS AND DISCUSSION

We have sequenced the genome of *L. edodes* L54A, a parental monokaryon of L54, by Roche 454 GS-FLX/Titanium and ABI SOLiD sequencing at over 11-fold coverage. The 39.8Mb draft genome sequence consists of 767 scaffolds with N50 sequence size of 111.9kb. Over 13,000 gene models were predicted from the genome sequence. Using the genome sequence of *L. edodes* L54A and 14 other fungal genomes, we have built a searchable genome database and analysis platform for bioinformatics and genome analysis.

To identify proteins conserved in all 4 Agaricales, predicted proteomes from *L. edodes* and 14 other fungal genomes were pooled and clustered to identify protein families. Over 1000 families are conserved among Agaricales but absent in ascomycetes (may present in at least one non-Agaricale basidiomycetes). To determine any enriched biological processes or molecular functions among the Agaricales-conserved protein families, Gene ontology (GO) terms of the corresponding *L. edodes* proteins were compared with that of all *L. edodes* proteins. A total of 54 GO Biological Process terms (Table 2) and 46 Molecular Function terms (Table 3) are significantly enriched ($P < 0.05$). Agaricale-conserved protein families were identified to be rich in putative regulators of biological processes, including gene expression (GO:0010468; P=1E-26), transcription (GO:0006355; P=3E-24), signaling (GO:0023051; P=9E-07). These regulators may represent a complex regulatory network in orchestrating fruiting body development of Agaricales. As there are Agaricales-conserved proteins enriched for nucleic acid binding transcription factor activity (GO:0001071; P=1E-15), these putative transcription factors could represent the switches in different stages of fruiting body development.

Table 2: Top 20 Gene Ontology (GO) Biological Process terms enriched among *L. edodes* proteins that are conserved among Agaricales but absent in ascomycetes

Accession	Term	Bonferroni corrected *P*-value	No. of *L. edodes* proteins
GO:0050789	regulation of biological process	2E-38	69
GO:0065007	biological regulation	3E-38	69
GO:0050794	regulation of cellular process	3E-37	67
GO:0019222	regulation of metabolic process	4E-27	47
GO:0010468	regulation of gene expression	1E-26	44
GO:0009987	cellular process	4E-26	157
GO:0060255	regulation of macromolecule metabolic process	4E-26	44
GO:0031323	regulation of cellular metabolic process	6E-26	45
GO:0080090	regulation of primary metabolic process	6E-26	45
GO:0019219	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	2E-25	44
GO:0051171	regulation of nitrogen compound metabolic process	2E-25	44
GO:0009889	regulation of biosynthetic process	7E-25	42
GO:0010556	regulation of macromolecule biosynthetic process	7E-25	42
GO:0031326	regulation of cellular biosynthetic process	7E-25	42
GO:2000112	regulation of cellular macromolecule biosynthetic process	7E-25	42
GO:0006355	regulation of transcription, DNA-dependent	3E-24	41
GO:0051252	regulation of RNA metabolic process	3E-24	41
GO:0044238	primary metabolic process	2E-16	120
GO:0043170	macromolecule metabolic process	4E-16	99
GO:0019538	protein metabolic process	7E-14	64

Table 3: Top 20 Gene Ontology (GO) Molecular Function terms enriched among *L. edodes* proteins that are conserved among Agaricales but absent in ascomycetes.

Accession	Term	Bonferroni corrected *P*-value	No. of *L. edodes* proteins
GO:0005488	binding	3E-59	226
GO:0005515	protein binding	1E-32	76
GO:0001071	nucleic acid binding transcription factor activity	1E-15	25
GO:0003700	sequence-specific DNA binding transcription factor activity	1E-15	25
GO:0046914	transition metal ion binding	5E-15	57
GO:0008270	zinc ion binding	2E-14	42
GO:0003824	catalytic activity	4E-14	156
GO:0046872	metal ion binding	6E-14	59
GO:0043167	ion binding	1E-13	59
GO:0043169	cation binding	1E-13	59
GO:0003676	nucleic acid binding	1E-13	74
GO:0003677	DNA binding	4E-09	44
GO:0030695	GTPase regulator activity	3E-08	14
GO:0004672	protein kinase activity	4E-08	26
GO:0060589	nucleoside-triphosphatase regulator activity	5E-08	14
GO:0005085	guanyl-nucleotide exchange factor activity	7E-08	11
GO:0016301	kinase activity	9E-08	28
GO:0030234	enzyme regulator activity	2E-07	15
GO:0016773	phosphotransferase activity, alcohol group as acceptor	3E-07	27
GO:0016787	hydrolase activity	7E-07	61

We have been using a battery of molecular techniques, including RNA arbitrarily primed-PCR, serial analysis of gene expression (SAGE), LongSAGE, cDNA sequencing and cDNA microarray, to analyze genes differentially expressed along the developmental stages [21, 22]. From mycelium to sporulating fruiting bodies, many physiological and biochemical changes occurs as revealed by analysis of the transcriptome. Through the transition from mycelial to primordial stages, different hydrophobins were expressed abundantly in the two stages, fewer structural genes were expressed, transcription and translation became active, and different genes involved in intracellular trafficking and stress responses were expressed. Massive cDNA pyrosequencing of mature fruiting bodies indicated that the mushroom (1) senses the external environment, (2) transmits signals to express genes through regulatory systems, (3) produces many proteins, (4) degrades unwanted proteins, (5) performs extensive biosynthesis, (6) generates energy, (7) regulates the internal environment, (8) transports molecules, (9) carries out cell division, and (10) differentiates and develops.

CONCLUSION

The genome sequence of *L. edodes* enriched genomic resources for mushroom research. Our works provided a holistic understanding of the molecular basis of the growth and development of mushrooms. Having a genome analysis platform, the mushroom community should also benefit from the access to the datasets in an organized manner.

ACKNOWLEDGEMENTS

This project was supported by the RGC General Research Fund CUHK467810 from the Research Grants Council of the HKSAR, PR China.

REFERENCES.

[1] Martin F. et al. (2008). The genome of *Laccaria bicolour* provides insight into mycorrhizal symbiosis. *Nature* 45: 88-92.
[2] Stajich J. E. et al. (2010). Insights into evolution of multicellular fungi from the assembled chromosomes of the mushroom *Coprinopsis cinerea (Coprinus cinereus*). *Proc. Natl. Acad. Sci. U.S.A.* 107: 11889-11894.
[3] Ohm R. A. et al. (2010). Genome sequence of the model mushroom *Schizophyllum commune. Nat. Biotechnol.* 28: 957-963.
[4] Stanke M. & Waack S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* 19: ii215-ii225.
[5] Zdobnov E.M. & Apweiler R. (2001). InterProScan--an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17: 847-848.
[6] Moriya Y. et al. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35: W182-W185.
[7] Hunter S. et al. (2009). InterPro: the integrative protein signature database. *Nucleic Acids Res.* 37: D211-D215.
[8] Altschul S. F. et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389-3402.
[9] Li L. et al. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13: 2178-2189.
[10] Loftus B. J. et al. (2005). The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans. Science* 307: 1321-1324.
[11] Martinez D. et al. (2004). Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. *Nat. Biotechnol.* 22: 695-700.
[12] Martinez D. et al. (2009). Genome, transcriptome, and secretome analysis of wood decay fungus *Postia placenta* supports unique mechanisms of lignocellulose conversion. *Proc. Natl. Acad. Sci. U. S. A.* 106: 1954-1959.
[13] Kämper J. et al. (2006). Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis. Nature* 444: 97-101.
[14] Nierman W. C. et al. (2005). Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus. Nature* 438: 1151-1156.
[15] Galagan J. E. et al. (2005). Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae. Nature* 438: 1105-1115.
[16] Machida M. et al. (2005). Genome sequencing and analysis of *Aspergillus oryzae. Nature* 438: 1157-1161.

[17] Galagan J. E et al. (2003). The genome sequence of the filamentous fungus *Neurospora crassa. Nature* 422: 859-868.
[18] Goffeau A. et al. (1996). Life with 6000 genes. *Science* 274: 546, 563-567.
[19] Martinez D et al. (2008). Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nat. Biotechnol.* 26: 553-560.
[20] Martin F. et al. (2010). Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature* 464: 1033-1038.
[21] Chum W. W. Y. et al. (2008). Gene expression studies of the dikaryotic mycelium and primordium of *Lentinula edodes* by serial analysis of gene expression. *Mycol. Research.* 112: 950-964.
[22] Chum W. W. Y. et al. (2011). Cataloging and profiling genes expressed in *Lentinula edodes* fruiting body by massive cDNA pyrosequencing and LongSAGE. *Fungal Genet. Biol.* 48: 359-369.